

## Дискуссии

© РОЗЕНТАЛЬ О.М., АЛЕКСАНДРОВСКАЯ Л.Н., 2019

Розенталь О.М.<sup>1</sup>, Александровская Л.Н.<sup>2</sup>

### РИСК-ОРИЕНТИРОВАННЫЙ ПОДХОД К ОЦЕНКЕ КАЧЕСТВА ВОДЫ ИСТОЧНИКОВ ПИТЬЕВОГО ВОДОСНАБЖЕНИЯ

<sup>1</sup> ФГБУН «Институт водных проблем» РАН, 119991, Москва;

<sup>2</sup> ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)», 125993, Москва

**Введение.** Корректный контроль и оценка соответствия состава и свойств воды источников питьевого водоснабжения установленным критериям безопасности являются необходимыми условиями обеспечения качества жизни и здоровья населения в ситуации нарастающего воздействия химических факторов. Поэтому важнейшее значение приобретает повышение достоверности санитарно-гигиенических заключений.

**Материал и методы.** Для достижения цели работы в ней использован байесовский подход, методы проверки однородности данных и оценка меры статистической однородности. Показано, что оценивание по методу Байеса, требуя учета ограничивающих факторов, обеспечивает корректное решение задачи о выполнении установленных санитарно-гигиенических требований. Представление байесовских соотношений в виде оценок смеси распределений позволяет учесть степень статистической однородности объединяемых выборок, что существенно расширяет область применения метода.

**Результаты.** На примере исследования качества воды Верх-Исетского водохранилища – рекреационной зоны и источника водоснабжения северной части г. Екатеринбурга показано, что учет степени однородности объединяемых данных позволяет «забывать» более раннюю информацию (в примере – за 2009 год) и тем самым придавать больший вес свежим данным (2010 года). Также показано, что в данном случае объединенные оценки с учетом степени однородности рядов данных оказываются не хуже байесовских оценок. Сделан вывод о том, что чем ниже степень однородности, тем быстрее происходит вышеуказанное «забывание». А риски нарушения гигиенических требований зависят от выбора закона распределения контролируемых показателей.

**Заключение.** Установлена возможность повышения достоверности санитарно-гигиенических заключений путем корректного объединения свежей измерительной информации с устаревшими данными с учетом их постепенного «забывания». Разработанные методы использованного для этого риск-ориентированного подхода могут найти широкое применение для оценки гигиенической безопасности населения в условиях воздействия широкого круга химических факторов.

**Ключевые слова:** гигиеническая безопасность населения в условиях воздействия химических факторов; оценка риска для здоровья; выборочный контроль; статистическая однородность; метод Байеса; объединение априорной и вновь полученной информации.

**Для цитирования:** Розенталь О.М., Александровская Л.Н. Риск-ориентированный подход к оценке качества воды источников питьевого водоснабжения. *Гигиена и санитария*. 2019; 98(5): 563-569. DOI: <http://dx.doi.org/10.18821/0016-9900-2019-98-5-563-569>

**Для корреспонденции:** Розенталь Олег Моисеевич, доктор тех. наук, главный научный сотрудник ФГБУН «Институт водных проблем» РАН, 119991, г. Москва. E-mail: [orosental@gambler.ru](mailto:orosental@gambler.ru)

**Финансирование.** Исследование не имело спонсорской поддержки.  
**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила 14.05.2018  
Принята к печати 06.02.2019  
Опубликована 06.2019

Rosenthal O.M.<sup>1</sup>, Aleksandrovskaya L.N.<sup>2</sup>

### RISK-ORIENTED APPROACH TO THE QUALITY ASSESSMENT OF WATER SOURCES OF DRINKING WATER SUPPLY

<sup>1</sup>Institute of Water Problems of RAS, Moscow, 119991, Russian Federation;

<sup>2</sup>Moscow Aviation Institute, Moscow, 125993, Russian Federation

**Introduction.** Correct control and assessment of compliance of the composition and properties of drinking water sources with the established safety criteria are necessary conditions for ensuring the quality of life and health of the population in a situation of increasing exposure to chemical factors. Therefore, the most important is to increase the reliability of sanitary and hygienic conclusions.

**Material and methods.** To achieve the goal of the work, there was used the Bayesian approach, methods for checking data homogeneity and evaluation of the measure of the statistical homogeneity. It is shown that the evaluation by Bayes method, requiring consideration of limiting factors, provides a correct solution to the problem of the implementation of the established sanitary and hygienic requirements. Representation of Bayesian relations in the form of estimates of the mixture of distributions allows taking into account the degree of statistical homogeneity of the combined samples, which significantly expands the scope of the method. Although the samples selected for consideration in the simple case can be considered homogeneous in the examples of the assessment of hygienic safety of water, it turned out that the transition from the binomial distribution law to a more strict normal one reduces the risk of accidental error of

the desired conclusion. This provides the researcher with the opportunity to use, depending on the requirements for the accuracy of the assessment, different algorithms of a risk-based approach to the assessment of hygienic safety.

**Results.** On the example of the study of the water quality in the Verkhny Isetsy Vodokanal recreational area and source of water in Northern part of Ekaterinburg the inclusion of the homogeneity of merged data was shown to allow “forgetting” the earlier information (e.g., 2009) and thus to give more weight to recent data (2010). It is also shown that in this case, the combined estimates with the degree of homogeneity of the data series are not worse than the Bayesian estimates. It is concluded that the lower the degree of uniformity, the faster the above “forgetting” occurs. And the risks of violation of hygienic requirements depend on the choice of the law of distribution of controlled indicators.

**Conclusion.** There was established the possibility of increasing the reliability of sanitary and hygienic conclusions by correctly combining the latest measurement information with outdated data, taking into account their gradual “forgetting”. The developed methods of the used for this risk-based approach can be widely applied to assess the hygienic safety of the population under the influence of a wide range of chemical factors.

**Key words:** hygienic safety of the population in terms of exposure to chemical agents; evaluation of health risk; sampling; statistical homogeneity; the Bayesian method; combining a priori and newly acquired information.

**For citation:** Rosenthal O.M., Aleksandrovskaya L.N. Risk-oriented approach to the quality assessment of water sources of drinking water supply. *Gigiena i Sanitariya (Hygiene and Sanitation, Russian journal)* 2019; 98(5): 563-569. (In Russ.). DOI: <http://dx.doi.org/10.18821/0016-9900-2019-98-5-563-569>

**For correspondence:** Oleg M. Rosenthal, MD, Ph.D., DSci., chief researcher of the Institute of water problems RAS, Moscow, 119991, Russian Federation. E-mail: [orosental@rambler.ru](mailto:orosental@rambler.ru)

**Information about the author:** Rosenthal O. M., <https://orcid.org/0000-0001-6261-6060>

**Conflict of interest.** The authors declare no conflict of interest.

**Acknowledgments.** The study had no sponsorship.

Received: 14 May 2018

Accepted: 06 February 2019

Published 06.2019

## Введение

Количество химических соединений, попадающих в реестр Chemical Abstracts Service за последнее десятилетие, ежедневно увеличивается на 50 000. Таковы данные идентификатора CAS registry number, обостряющие задачу обеспечения гигиенической безопасности населения в условиях накопления старых и появления новых веществ и материалов [1]. Возрастает актуальность доказательной оценки неблагоприятного химического воздействия [2, 3], ориентирующая органы санитарно-эпидемиологического надзора на ограничение соответствующего риска приемлемым уровнем. Решение поставленной задачи требует достоверной информации о содержании и свойствах веществ, загрязняющих объекты окружающей среды [3–6]. К сожалению, получение такой информации сопровождается трудностями, обусловленными погрешностью выборочных измерений и нестабильностью (вариабельностью [7]) показателей, например, состава пресной воды [8], особенно быстро меняющегося в промышленных регионах с повышенным уровнем водопотребления и водоотведения [9, 10]. Поэтому не только санитарно-гигиеническая оценка воздействия химических соединений, но и информация об их концентрации должны отвечать требованию приемлемости риска ошибки.

Пусть, например, решается задача обеспечения гигиенической безопасности источника питьевого водоснабжения [11–13] по результатам исследования концентрации одного из загрязняющих воду веществ. Если оценка вероятности выполнения установленных требований  $\bar{R} = \frac{m}{n}$ , то тогда риск их нарушения  $\bar{r} = \frac{d}{n}$  зависит от общего количества измерений  $n$ , в том числе удовлетворительных  $m$  и неудовлетворительных  $d = n - m$ . Если приемлемый риск не превышает 0,05 и  $d = 1$ , то потребуется не менее  $n = \frac{1}{0,05} = 20$  измерений. Отсюда следует, что при проведении плановых измерений контролируемых показателей четыре раза в год необходимо использовать данные за пять лет, а если ежемесячно, то почти за два года. Простое объединение данных за столь длительные сроки нежелательно, поскольку качество воды за такой длительный промежуток времени может заметно измениться, например, из-за непостоянства экономических или экологических факторов региона.

Для объединения информации, учитывающего её постепенное «устаревание», ниже предлагается методика оценки прием-

лемого риска  $\bar{r}$ , в основу которой положен байесовский подход объединения априорной и вновь полученной информации [14].

## Материал и методы

Формула Байеса для совместной вероятности событий  $A$  и  $B$  имеет вид:

$$P(A, B) = P(A|B)P(B) = P(B|A)P(A)$$

или

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \quad (1)$$

где  $P(B)$  – априорная плотность вероятности оцениваемого параметра;  $P(B|A)$  – апостериорная плотность вероятности;  $P(A|B)$  – функция правдоподобия, представляющая собой плотность распределения вероятностей данных при фиксированных параметрах, определяющих эту плотность;  $P(A) = \int_{\Omega} P(A|B)P(B)dB$  – безусловная плотность вероятности события  $A$  (нормирующий множитель)  $\Omega$  – область определения параметра  $B$ .

В математической статистике хорошо известны оценки параметров распределений, получаемые из условия максимума функции правдоподобия [15]. При байесовском подходе функции распределения этих оценок уточняются путём учёта информации, имеющейся до получения очередных данных измерений [16–21].

В простейшем случае байесовское оценивание осуществляется путём оценки вероятности по частоте. Тогда функция правдоподобия представляет собой биномиальное распределение

$$P(m|n, R_n) = \frac{m!}{n!(n-m)!} R_n^m (1-R_n)^{n-m}, \quad (2)$$

где  $R_n$  – оценка вероятности соответствия контролируемого показателя установленным гигиеническим требованиям; индекс «n» здесь и далее характеризует правдоподобие величины.

Априорную плотность вероятности рекомендуется задавать так, чтобы она была [22]: «самовоспроизводимой» (после коррекции по Байесу не должен меняться вид распределения), многообразной по форме (для адекватного описания априорных данных), «наихудшей» для гарантии точности получаемых априорных оценок (этому требованию отвечают распределения, полученные из условия максимальной неопределенности).

Априорные и апостериорные распределения для нормального распределения

Функция правдоподобия	Априорное распределение	Апостериорное распределение
$P(x_1, \dots, x_n / m_x, h_x) = (2\pi)^{-\frac{1}{2}n} \times$ $\times \exp\left\{-\frac{1}{2}h_x(n-1)S^2 - \frac{1}{2}h_x n(\bar{x} - m_x)^2\right\} \times h_x^{\frac{1}{2}n},$ <p>где <math>\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i</math>; <math>S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2</math> (<math>n &gt; 1</math>);  <math>S^2 = 0</math> (<math>n = 1</math>); <math>h_x = \frac{1}{S^2}</math>.</p>	<p>Гамма-нормальное распределение</p> $P(m_0, h_0) = (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}h_0 n_0 (m_0 - \bar{x}_0)^2\right\} \times (h_0 n_0)^{\frac{1}{2}} \times$ $\times \exp\left\{-\frac{1}{2}h_0 v_0 S_0^2\right\} \left(\frac{1}{2}h_0 v_0 S_0^2\right)^{\frac{1}{2}v_0-1} \times \frac{\frac{1}{2}v_0 S_0^2}{\left(\frac{1}{2}v_0 - 1\right)!}$ <p>с параметрами: <math>\bar{x}_0, n_0, S_0^2, v_0 = n_0 - 1, h_x = \frac{1}{S_0^2}</math>.</p>	<p>Гамма-нормальное совместное распределение:</p> $\bar{x}_6 = \frac{\bar{x}_0 n_0 + \bar{x} n}{n_0 + n};$ $n_6 = n_0 + n; v_6 = n_6 - 1;$ $S_6^2 = \frac{S_0^2 (n_0 - 1) + S^2 (n - 1) + (\bar{x} - \bar{x}_0)^2 \frac{n_0 n}{n_0 + n}}{n_0 + n - 1}$ <p>Одномерные распределения:          для <math>m_6 - t</math>-распределение Стьюдента:</p> $P(m_6 / x_1, \dots, x_n) = \frac{v_6^{1/2} v_6}{B\left(\frac{1}{2}, \frac{1}{2} v_6\right)} \times$ $\times \left[ v_6 + \frac{(m_6 - \bar{x}_6)^2}{S_6^2 / \sqrt{n_6}} \right]^{-\frac{1}{2}v_6-1} \frac{\sqrt{n_6}}{S_6^2}$ <p>с математическим ожиданием <math>M[m_6] = \bar{x}_6</math>          и дисперсией <math>D[m_6] = (S_6^2 / n_6) \cdot (v_6 / v_6 - 2)</math>;</p> <p>для <math>h_6 -</math> Гамма-распределение второго рода:</p> $P(h_6 / x_1, \dots, x_n) = \frac{1}{\left(\frac{1}{2}v_6 - 1\right)!} \exp\left\{-\frac{1}{2}v_6 h_6 S_6^2\right\} \times$ $\left(\frac{1}{2}v_6 h_6 S_6^2\right)^{\frac{1}{2}v_6-1} \frac{1}{2}v_6 S_6^2$ <p>с математическим ожиданием <math>M[h_6] = \frac{1}{S_6^2}</math>          и дисперсией <math>D[h_6] = \frac{1}{\frac{1}{2}v_6 (S_6^2)^2}</math> или <math>M[\sigma_6^2] = S_6^2</math>          и <math>D[\sigma_6^2] = \frac{S_6^2}{v_6}</math></p>

Для функции (2) перечисленным условиям удовлетворяет бета-распределение [23]:

$$P(R_0) = \beta(\gamma_0, \eta_0) R_0^{\gamma_0-1} (1 - R_0)^{\eta_0-1}, \tag{3}$$

где  $\beta(\gamma_0, \eta_0)$  – бета-функция;  $\gamma_0 - 1, \eta_0 \geq 1$  – её параметры; индекс «0» здесь и далее относит величину к априорному распределению.

Подставляя (2) и (3) в (1) и вводя новые параметры, получим выражение для апостериорной вероятности:

$$P(R_6 | m, n) = \beta(\gamma, \eta) R_6^{\gamma-1} (1 - R_6)^{\eta-1}, \tag{4}$$

где  $R_6$  – байесовская оценка вероятности (индекс «б»).

Здесь условием применимости байесовского подхода является равенство истинных значений параметров  $R_0 = R_{\Pi} = R_6$ , тогда как при оценивании этих величин, в том числе на уровне допустимого риска, вполне возможно их неравенство:  $R_0 \neq R_{\Pi} \neq R_6$ .

В качестве оценки параметра  $R$  выбираем апостериорное математическое ожидание

$$\bar{R}_6 = M[R_6 | m, n] = \frac{\gamma}{\gamma + \eta}. \tag{5}$$

Если параметры  $\gamma_0, \eta_0$  получены по результатам предшествующих измерений, то  $\gamma_0 = m_0; \eta_0 = n_0 - m_0$ . При этом  $\bar{R}_0 = \frac{m_0}{n_0}$  – оценка вероятности по частоте, найденная в предыдущий период. Для этого случая

$$\bar{R}_6 = \frac{m_0 + m}{n_0 + n} = \frac{n_0}{n_0 + n} \bar{R}_0 + \frac{n}{n_0 + n} \bar{R}_n, \tag{6}$$

т. е. байесовской является оценка по объединённой выборке. В [24] показано, что этот вывод справедлив и для других распределений.

Подобная оценка частоты появления соответствия/несоответствия контролируемых показателей гигиеническим требованиям универсальна, но недостаточно информативна. Возможно повышение точности вывода путём использования более «сильной» метрологической шкалы, чем номинальная, опирающаяся на биномиальный закон распределения. Это усложняет соответствующие вычисления, но оправдано в условиях повышающихся требований к безопасности населения при воздействии химических факторов.

В таблице приведены априорные и апостериорные распределения вероятностей для функции правдоподобия – нормально-гамма распределения с неизвестными математическим ожиданием и дисперсией. Нетрудно показать, что и здесь полученные байесовские оценки математического ожидания  $\bar{x}_6$  и дисперсии  $S_6^2$  есть не что иное, как оценки по объединённой выборке<sup>1</sup>.

<sup>1</sup> Подобные совпадения справедливы только если предварительная и вновь полученная информация оцениваются в одинаковых шкалах измерений; в общем же случае байесовский подход имеет более широкую область применения, поскольку в качестве априорной информации могут выступать экспертные оценки, гигиенические исследования объектов-аналогов, данные формального моделирования и т. д. [24, 30].

Обозначим оцениваемый параметр в формулах (7), (9), (10) через  $\Theta$  (это может быть математическое ожидание, дисперсия и др.). Тогда эти оценки могут быть представлены в виде:

$$\bar{\Theta}_6 = p_1 \bar{\Theta}_0 + p_2 \bar{\Theta}_n = \frac{n_0}{n_0 + n} \bar{\Theta}_0 + \left(1 - \frac{n_0}{n_0 + n}\right) \bar{\Theta}_n. \quad (7)$$

При этом условием несмещённости (отсутствия систематической погрешности) этих оценок является принадлежность объединяемых выборок к одной генеральной совокупности (статистической однородности данных). Действительно, тогда  $M[\bar{\Theta}_6] = p_1 M[\bar{\Theta}_0] + p_2 M[\bar{\Theta}_n] = (p_1 + p_2) M[\bar{\Theta}] = \Theta$ . Поэтому процедуре объединения данных должна предшествовать проверка статистической однородности, т. е. равенства параметров соответствующих распределений.

Если имеется неравенство  $R_0 \neq R_n$ , то вероятность события  $(d_0, d)$  представляет собой совместную вероятность двух независимых событий  $d_0$  и  $d$ , обозначающих числа отказов, зафиксированных при предварительных и последующих наблюдениях:

$$p(d_0, d) = \frac{n_0!}{d_0!(n_0 - d_0)!} R_0^{n_0 - d_0} (1 - R_0)^{d_0} \frac{n!}{d!(n - d)!} R_n^{n - d} (1 - R_n)^d$$

При  $R_0 = R_n = R_6$  две выборки объёмов  $n_0$  и  $n$  могут быть объединены, что приводит к совместной вероятности вида:

$$p(d_0 + d) = \frac{(n_0 + n)!}{(d_0 + d)![(n_0 + n) - (d_0 + d)]!} R^{(n_0 + n) - (d_0 + d)} (1 - R)^{d_0 + d}$$

Отношение этих вероятностей подчиняется гипергеометрическому распределению:

$$\frac{p(d_0, d)}{p(d_0 + d)} = \frac{\binom{n_0}{d_0} \binom{n}{d}}{\binom{n_0 + n}{d_0 + d}}$$

где  $\binom{n_0}{d_0} = \frac{n_0!}{d_0!(n_0 - d_0)!}$ ;  $\binom{n}{d} = \frac{n!}{d!(n - d)!}$ ;

$$\binom{n_0 + n}{d_0 + d} = \frac{(n_0 + n)!}{(d_0 + d)![(n_0 + n) - (d_0 + d)]!}$$

Областью принятия нулевой гипотезы  $R_0 = R_n = R_6$  здесь является область, ограниченная условием:

$$p(q \geq d) = \sum_{q=0}^n \binom{n}{q} \binom{n_0}{D - q} / \binom{n_0 + n}{D} = \frac{\alpha}{2} \text{ при альтернативной гипотезе } R_0 > R_n, \text{ где } D = d_0 + d; \alpha - \text{выбранный уровень значимости.}$$

Используя биномиальную аппроксимацию гипергеометрического распределения, получим

$$\sum_{q=d}^{d_0 + d} \binom{d_0 + d}{q} \left(\frac{n}{n_0 + n}\right)^q \left(1 - \frac{n}{n_0 + n}\right)^{d_0 + d - q} = \frac{\alpha}{2},$$

откуда при выборе  $d = 1$  следует простое условие, связывающее объемы выборки  $n_0$  и число испытаний  $n$  до первого отказа:

$$1 - \left(1 - \frac{n}{n_0 + n}\right)^{d_0 + 1} = \frac{\alpha}{2}. \quad (8)$$

Если  $d_0 = 0$  и  $n_0 = n$ , то тогда  $\alpha = 1$ , т. е. при отсутствии несоответствий выборки равных объёмов всегда будут признаны однородными. Аналогичный результат может быть получен, если рассмотреть вероятности не числа несоответствий установленным требованиям, а числа соответствий. Случай отсутствия соответствия особенно важен, если по ряду загрязняющих веществ наблюдается постоянное превышение ПДК.

Необходимым условием равенства параметров нормального закона распределения является выполнение неравенства Фишера:

$$F_0 = \frac{S_0^2}{S^2} \leq F_{1-\alpha}(n_0 - 1, n - 1) \text{ при } S_0^2 > S^2$$

или

$$F_0 = \frac{S^2}{S_0^2} \leq F_{1-\alpha}(n - 1, n_0 - 1) \text{ при } S^2 > S_0^2$$

где  $F_{1-\alpha}$  – квантиль уровня  $1-\alpha$  распределения Фишера при указанном в скобках числе степеней свободы.

При принятии решения о равенстве дисперсий критерием проверки равенства математических ожиданий является требование выполнения условия:

$$t_{01} = \frac{|\bar{x}_0 - \bar{x}|}{S' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2),$$

где  $(S')^2 = \frac{S_0^2(n - 1) + S^2(n - 1)}{n_1 + n_2 - 1}$  – объединённая дисперсия

для гипотезы равенства математических ожиданий;  $t_{1-\frac{\alpha}{2}}$  – квантиль уровня  $1-\alpha/2$  распределения Стьюдента с  $n_1 + n_2 - 2$  числом степеней свободы.

При решении о неравенстве дисперсий используется приближённый критерий Саттервейта: [25, 26]:

$$t_{02} = \frac{|\bar{x}_0 - \bar{x}|}{\sqrt{\frac{S_0^2}{n_1} + \frac{S^2}{n_2}}} \leq t_{1-\frac{\alpha}{2}}(v),$$

где число степеней свободы  $v = \frac{(S_0^2/n_0 + S^2/n)^2}{\frac{(S_0^2/n_0)^2}{n_0 - 1} + \frac{(S^2/n)^2}{n - 1}}$  в зависи-

мости от конкретных значений  $S_0^2, S^2$ , меняется в диапазоне  $\min\{n_0 - 1, n - 1\} \leq v \leq (n_0 + n - 2)$ .

При проверке равенства параметров нормального распределения критерий Саттервейта может рассматриваться как комплексный, т. к. он не требует предварительной проверки равенства дисперсий, а возможное их неравенство отражается в изменении числа степеней свободы. Действительно, пусть  $n_0 = n$  и  $S_0^2 = S^2$ . Тогда  $v = 2(n - 1)$  достигает своего максимально возможного значения. А при  $n_0 = n$  и наибольшей степени различия между оценками дисперсий, например,  $S_0^2 \approx 0 : v = n - 1$ , достигает своего минимального значения. В промежуточных случаях  $\min\{n_0 - 1, n - 1\} \leq v \leq (n_0 + n - 2)$ , причём при

$$n_0 = n \text{ и } \frac{S_0^2}{S^2} = F_{1-\alpha}(n_0 - 1, n - 1), v = \frac{(n - 1)(1 + F_{1-\alpha})^2}{1 + F_{1-\alpha}^2}.$$

Так, при  $n_0 = n = 12, \alpha = 0,05, F_{1-0,05} = 2,82; v = 1,63(n - 1) \approx 18$  (округление до целого числа), т. е. лежит в интервале между 11 и 22.

Как при проверке любой статистической гипотезы, принимаемое решение об однородности/неоднородности связано с ошибками. Используя терминологию допускового контроля, можно считать, что если исследуемая статистика  $(F_0, t_{01}, t_{02})$  находится в допуске, ограниченном квантилями соответствующих распределений, то с вероятностью  $1-\alpha$  принимается решение об однородности. Однако достоверность оценок зависит от расстояния статистики до границы поля допуска [27, 28]. Это расстояние в вероятностной форме представляет собой наблюдаемый уровень значимости  $\alpha_p$ , найденный из условия точного равенства значения статистики квантилю распределения.

Границы допуска обычно определяются на практике при значениях  $\alpha_p = [0,05 \div 1]$ . А для использования  $\alpha_0$  в качестве «меры» однородности необходимо нормирование  $\alpha^* = \frac{\alpha_0 - \alpha_{п}}{1 - \alpha_{п}}$ . Тогда при  $\alpha_0 = 1 : \alpha^* = 1$ , при  $\alpha_0 = \alpha_{п} : \alpha^* = 0$ . Только при уменьшении граничного значения  $\alpha_{п} \rightarrow 0 : \alpha^* \rightarrow 0$  нормирования не требуется.

Таблица 2

Концентрация загрязняющих воду Верх-Исетского водохранилища веществ в 2010 и 2009 гг. при ежемесячных измерениях на створе водозабора ТЭЦ

Химический элемент	Год	Концентрация, мг/дм <sup>3</sup>											
		0,11	0,56	1,18	0,15	0,17	0,82	0,16	1,39	1,69	0,78	1,09	0,97
Железо общее	2010												
Марганец		0,13	0,05	0,36	0,38	0,47	0,42	0,37	0,43	0,48	0,39	0,41	0,46
Железо общее	2009	0,45	0,36	1,39	0,33	2,12	2,19	0,96	1,54	0,32	0,52	0,31	0,38
Марганец		0,06	0,49	0,38	0,57	0,21	0,04	0,78	0,09	0,41	0,03	0,11	0,15

### Использование предлагаемого подхода к оценке качества воды источников питьевого водоснабжения

**Пример 1.** Для Верх-Исетского водохранилища – источника питьевого водоснабжения северной части г. Екатеринбурга – необходимо оценить наблюдаемый уровень значимости принятия решения об однородности рядов данных концентрации железа и меди – веществ, из-за избытка которых страдают внутренние органы жителей города и возникают онкологические заболевания.

**Решение.** Как видно из табл. 2 (строки 2 и 3), концентрация (здесь и далее выраженная в мг/дм<sup>3</sup>) в 2010 г. загрязняющих воду веществ изменяется в широких пределах и лишь частично соответствует гигиеническим требованиям, равным 0,3 для железа и 0,5<sup>2</sup> для марганца. Сравним эти данные с априорными 2009 г. (см. табл. 2 строки 4, 5).

Из табл. 3 видно, что в 2010 году по сравнению с 2009 г. среднеарифметическая концентрация загрязняющих веществ в воде понизилась на 15–20%; значительные изменения характерны и для дисперсии, вследствие чего решение об однородности/неоднородности не может быть однозначным.

Для марганца число несоответствий гигиеническим нормативам в 2009 г. равно  $d_0 = 2$ , в 2010 г. –  $d = 0$ . При выборе в качестве модели биномиального распределения обе выборки могут быть признаны однородными с  $\alpha_0 = 0,25$ . То же имеем для железа уже с  $\alpha_0 = 0,15$ , поскольку здесь число соответствий  $m_0 = 0$ ;  $m = 4$ . При использовании нормального распределения имеем другие данные (см. табл. 3). И хотя во всех рассмотренных случаях  $\alpha_0 > \alpha_{cr}$ , то всё же видно, что при более строгом подходе с заменой биномиального распределения вероятностей нормальным, уровень значимости принятия решения об однородности для марганца снижается почти вдвое, приближаясь к границе допуска, а для железа втрое возрастает, повышая гарантию сохранения однородности.

### Расширение возможностей предлагаемого подхода с использованием обобщённой формы объединённых оценок

Для обобщения байесовского подхода с учётом возможной степени статистической неоднородности объединяемых данных рассмотрим две гипотезы:

1. гипотезу однородности объединённых данных  $U$  с вероятностью  $P(U)$ ;
2. гипотезу неоднородности  $\bar{U}$  с вероятностью  $P(\bar{U}) = 1 - P(U)$ .

Тогда, в соответствии с формулой полной вероятности, апостериорная безусловная плотность вероятности из (4) запишется в виде  $P(R_{06} | m, n) = P(R_{06} | m, n, U)P(U) + P(R_{06} | m, n, \bar{U})P(\bar{U})$ , где индекс «об» означает объединённую оценку.

Если выполняется гипотеза однородности с  $U(P|U) = 1$ , то выражение  $P(R_{06} | m, n)$  есть байесовская апостериорная плотность вероятности; если же выполняется гипотеза неоднородности  $\bar{U}(P|\bar{U}) = 1$ , то целесообразно отказаться от использования априорной информации, и оценку производить только по апостериорным данным (оценка максимального правдоподобия

<sup>2</sup> Здесь и далее величина, указанная в скобках устанавливается по постановлению Главного государственного санитарного врача по соответствующей территории для конкретной системы водоснабжения

[29]). В результате переходим к модели так называемой смеси распределений [30], которая описывается формулой

$$P(x) = \sum_{j=1}^k p_j f_j(x, \Theta_j),$$

где  $\sum_{j=1}^k p_j = 1$ ;  $f_j(x, \Theta_j)$ ,  $f(x)$  в непрерывном случае – плотности, а в дискретном случае – ряды частот соответствующей  $j$ -той компоненты смеси и результирующего закона распределения;  $p_j$  – априорная вероятность появления в случайной выборке наблюдения с законом распределения  $f_j(x, \Theta_j)$  (т. е. удельный вес таких наблюдений в общей генеральной совокупности);  $k$  – число компонент смеси.

С законами распределения подобной структуры исследователь сталкивается, например, в ситуациях, когда ему приходится анализировать генеральную совокупность, объединяющую в себе несколько подсовкупностей, каждая из которых однородная (в смысле унимодальности соответствующего закона распределения  $f_j(x, \Theta_j)$ ), но существенно отличается от других, например, значением параметра  $\Theta_j$ . При этом параметр  $\Theta_j$  может определять как центр группирования соответствующих наблюдений (тогда он интерпретируется как параметр сдвига), так и их меру случайного рассеивания (тогда он интерпретируется как параметр масштаба).

Начальные моменты смеси распределений определяются по формуле:

$$a_r(x) = \sum_{j=1}^k p_j a_{rj},$$

где  $a_r(x)$  –  $r$ -тый начальный момент;  $p_j$  – доля  $j$ -го распределения в смеси;  $a_{rj}$  –  $r$ -тый начальный момент  $j$ -го распределения. Когда математические ожидания таких распределений совпадают, то центральные моменты определяются по формулам, аналогичным записанным для начальных моментов.

Таким образом, для построения обобщённой формы объединённых оценок с учётом степени возможной неоднородности используется смесь распределений с  $P(U) = \alpha^*$ :

$$P_{06}(\Theta) = \alpha^* P(\Theta_5) + (1 - \alpha^*) P(\Theta_n), \quad (9)$$

где при  $\alpha^* = 1 - P_{06}(\Theta) = P(\Theta_5)$  – апостериорная байесовская плотность вероятности; при  $\alpha^* = 0 - P_{06}(\Theta) = P(\Theta_n)$  – функ-

Таблица 3

### Проверка статистической однородности выборочных данных

Химический элемент	Год	Параметр	$\nu$	$t_{02}$	$\alpha_0$	
Mn	2009	Математическое ожидание $\bar{x}_0$	0,27	≈ 15	1,4	0,16
		Дисперсия $S_0^2$	0,04			
	2010	Математическое ожидание $\bar{x}_0$	0,91			
		Дисперсия $S_1^2$	0,01			
Fe	2009	Математическое ожидание $\bar{x}_0$	0,91	≈ 20	0,7	0,47
		Дисперсия $S_0^2$	0,36			
	2010	Математическое ожидание $\bar{x}_0$	0,75			
		Дисперсия $S^2$	0,19			

Таблица 4

Риск нарушения норматива

Химический элемент	Нормальное распределение						Биномиальное распределение	
	$\bar{x}_6$	$S_6$	$\bar{x}_{об}$	$S_{об}$	$\bar{r}_6$	$\bar{r}_{об}$	$r_6$	$r_{об}$
Mn	0,32	0,17	0,35	0,11	0,19	0,09	0,08	0,02
Fe	0,83	0,53	0,79	0,49	0,84	0,84	0,83	0,72

ция правдоподобия. Из (9) вытекает выражение для оценок  $\bar{\Theta}_{об} = \alpha^* \bar{\Theta}_6 + (1 - \alpha^*) \bar{\Theta}_{II}$  и с учетом (7):

$$\bar{\Theta}_{об} = \frac{n_0 \alpha^*}{n_0 + n} \bar{\Theta}_0 + (1 - \frac{n_0 \alpha^*}{n_0 + n}) \bar{\Theta}_{II} \quad (10)$$

Сравнивая это выражение с байесовской оценкой (6) приходим к выводу о перераспределении суммарного объёма двух выборок  $n_0 + n$  в сторону уменьшения «относительного веса» априорного объёма ( $\frac{n_0 \alpha^*}{n_0 + n}$  вместо  $\frac{n_0}{n_0 + n}$ ), т. е. «забывания» устаревшей информации, и увеличения «относительного веса» вновь поступающих данных –  $(1 - \frac{n_0 \alpha^*}{n_0 + n})$  вместо  $(1 - \frac{n_0}{n_0 + n})$ , тем более интенсивного, чем меньше степень однородности объединяемых данных.

Демонстрация оценки риска нарушения гигиенических требований

**Пример 2.** По результатам примера 1 требуется рассчитать байесовские оценки вероятности нарушения установленных требований без учёта и с учётом степени однородности объединяемых данных, т. е. соответственно риски  $r_6$  и  $r_{об}$ .

**Решение.** Результаты расчёта по (10) сведены в табл. 4.

Здесь оценки  $\bar{x} = \bar{x}_6, \bar{x}_{об}; S = S_6, S_{об}; \bar{r} = \bar{r}_6, \bar{r}_{об}$ , где риск оценивали для биномиального распределения по формуле  $\bar{r} = 1 - \frac{m}{n}$ ; для нормального – по  $\bar{r} = 1 - \Phi(\frac{ПДК - \bar{x}}{S})$ ;  $\Phi$  – функция стандартного нормального распределения.

Для большей наглядности выводу введём понятие эквивалентных биномиальных результатов измерений (индекс «эквб»), и представим оценки искомого риска в виде  $\bar{r} \approx \frac{d_{эквб,об}}{n_{эквб,об}}$  с округлением  $d_{эквб}$  до ближайшего целого числа (табл. 5).

Из табл. 5 следует, что риски нарушения гигиенических требований, оцененные с использованием различных законов распределения, значительно различаются. Для марганца это различие проявляется вдвое при использовании объединённой выборки, а в большинстве других случаев – на 10–30%. Это объясняется тем, что использование нормального закона распределения позволяет помимо прочего оценить близость контролируемых показателей к допустимым границам, что особенно важно при риск-ориентированном подходе к прогнозу качества воды. Что же касается учёта степени однородности объединяемых данных, то это позволяет «забывать» более раннюю информацию (2009 г.) и придавать больший вес свежим данным (2010 г.). В частности, в данном случае объединённые оценки с учётом степени однородности оказываются не хуже байесовских оценок, поскольку учитывают улучшение состояния воды в 2010 г. по сравнению с 2009 г. В общем же случае, чем ниже степень однородности, тем отчетливее выражено такое «забывание».

Заключение

Байесовское оценивание обеспечивает риск-ориентированный подход к оценке факторов гигиенической безопасности путём объединения накопленных за прошедшие периоды наблюдений, повышающих репрезентативность информации. Установлено, что требуя определённых усилий и учёта ограничивающих факторов, такое оценивание позволяет корректно

Таблица 5

Эквивалентные биномиальные результаты измерений

Химический элемент	Нормальное распределение				Биномиальное распределение			
	$r_0$	$r_{II}$	$r_6$	$r_{об}$	$r_0$	$r_{II}$	$r_6$	$r_{об}$
Mn	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{3}{24}$	$\frac{2}{24}$	$\frac{2}{12}$	$\frac{0}{12}$	$\frac{2}{24}$	$\frac{1}{24}$
Fe	$\frac{10}{12}$	$\frac{10}{12}$	$\frac{20}{24}$	$\frac{20}{24}$	$\frac{12}{12}$	$\frac{8}{12}$	$\frac{20}{24}$	$\frac{18}{24}$

решать задачи о выполнении установленных санитарно-гигиенических требований. Это показано на рассмотренном примере исследования качества воды Верх-Исетского водохранилища – рекреационной зоны и источника водоснабжения части г. Екатеринбург.

Представление байесовских соотношений в виде оценок смеси распределений обеспечивает учёт степени статистической однородности объединяемых выборок, что существенно расширяет область применения метода Байеса для гигиенических оценок, обеспечивая, при необходимости, повышенный уровень их достоверности. Например, выбор в качестве статистической модели вместо биномиального распределения нормального позволяет значительно уточнить требуемый вывод.

Приведённые примеры из области оценки качества воды источника питьевого водоснабжения носят демонстрационный характер. Они позволили показать возможность повышения достоверности санитарно-гигиенических заключений путём объединения измерительной информации при условии снижения значимости (постепенного «забывания») устаревающих данных. Разработанные методы использованного для этого риск-ориентированного подхода могут найти широкое применение для оценки гигиенической безопасности населения в условиях воздействия химических факторов.

Литература

(пп. 5, 8, 9, 14–22, 27–29 см. References)

1. Онищенко Г.Г., Рахманин Ю.А., Зайцева Н.В. и др. *Научно-методические аспекты обеспечения гигиенической безопасности населения в условиях воздействия химических факторов*. М.: МИГ «Медицинская книга», 2004. 368 с.
2. Новиков С.М., Фокин М.В., Унгурану Т.Н. Актуальные вопросы методологии и развития доказательной оценки риска здоровью населения при воздействии химических веществ. *Гигиена и санитария*. 2016; 95(8):711-716.
3. Р 2.1.10.1920-04. Руководство по оценке риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду. М.: Федеральный центр Госсанэпиднадзора России. 2004. 143 с.
4. Онищенко Г.Г., Новиков С.М., Рахманин Ю.А., Авалиани С.Л., Буштуева К.А. *Основы оценки риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду*. Под ред. Рахманина Ю.А., Онищенко Г.Г. М.: НИИ ЭЧ и ГОС. 2002. 408 с.
6. *Реки и озера мира. Энциклопедия*. Ред. В.И. Данилов-Данильян. Изд. Энциклопедия. М.: 2012. 925 с.
7. Ниворожкина Л.И., Морозова З.А. *Теория вероятностей и математическая статистика в определениях, формулах и таблицах: справочное пособие*. Ростов-н/Д: Феникс, 2007. 192с.
10. Розенталь О.М., Александровская Л.Н. *Водные ресурсы. Качество информации о составе воды*. 2015; 42 (4): 433-441.
11. СанПиН 2.1.5.980-00 «Гигиенические требования к охране поверхностных вод».
12. СанПиН 2.1.4.1074-01. Питьевая вода. Гигиенические требования к качеству воды централизованных систем питьевого водоснабжения. Контроль качества.
13. СанПиН 2.1.4.1175-02 Гигиенические требования к качеству воды нецентрализованного водоснабжения. Санитарная охрана источников.
23. Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. *Справочник по теории вероятностей и математической статистике*. М.: Наука, 1985. 640 с.

24. Райфа Г., Шлейфер Р. *Прикладная теория статистических решений*. М.: Статистика, 1977. 360 с.
25. ГОСТ Р 54500.3-2011/Руководство ИСО/МЭК 98-3:2008 Неопределенность измерения. Часть 3. Руководство по выражению неопределенности измерения.
26. ГОСТ Р ИСО 21748-2012. Статистические методы. Руководство по использованию оценок повторяемости, воспроизводимости и правильности при оценке неопределенности измерений.
30. Крюков С.П., Бодрунов С.Д., Александровская Л.Н. и др. *Методы анализа и оценивания рисков в задачах менеджмента безопасности сложных технических систем*. СПб.: Корпорация «Аэрокосмическое оборудование», 2007. 460 с.

## References

1. Onishchenko G.G., Rakhmanin Yu.A., Zajceva N.V. *Scientific and methodological aspects of ensuring the hygienic safety of the population in terms of exposure to chemical factors*. M.: MIG "Medical book", 2004. 368 p. (in Russian).
2. Novikov S. M., Fokin M. V., Ungureanu T. N. Topical issues of methodology and development of evidence-based assessment of health risk of the population under the influence of chemicals. *Gigiiena i sanitariya [Hygiene and Sanitation, Russian journal]*. 2016; 95(8): 711-716. (in Russian).
3. P 2.1.10.1920-04. Guidelines for the assessment of public health risks from exposure to chemicals that pollute the environment. M.: Federal center of Gossanepidnadzor of Russia. 2004. 143 p. (in Russian).
4. Onishchenko G.G., Novikov S.M., Rakhmanin Yu.A., Avaliani S.L., Bushueva K.A. *Principles of risk assessment for public health when exposed to chemicals, polluting the environment*. Ed. Rakhmanin Yu. A., Onishchenko G.G. M.: Institute of ECH and STATE. 2002. 408 p. (in Russian).
5. EPA (U.S. Environmental Protection Agency). 2007. Concepts, Methods, and Data Sources for Cumulative Health Risk Assessment of Multiple Chemicals, Exposures and Effects: a Resource Document. EPA/600/R-06/013F. August. 2007.
6. *Rivers and lakes of the world. Encyclopedia*. Ed. V. I. Danilov-Danilyan. Encyclopedia. M.: 2012. 925 p. (in Russian).
7. Nivorozhkina L. I., Morozova Z. A. *Probability Theory and mathematical statistics in definitions, formulas and tables: reference manual*. Rostov-on-Don: Phoenix, 2007. 192 p. (in Russian).
8. Kerry V. Smith, William H. *Measuring Water Quality Benefits*. Kluwer-Nijhoff Publishing. Boston, 2010. 326 p.
9. Rozental O. M., Chereshev V. A., A normal model of balanced nature management. *Russian Journal of Ecology*, 2010; 41(41): 356-363. © Pleiades Publishing, Ltd., 2010.
10. Rosenthal O. M., Alexander L. N. *Vodnye resursy. Kachestvo informacii o sostave vody [Water resources management. Quality of information on the composition of water]*. 2015; 42(4): 433-441. (in Russian).
11. SanPiN 2.1.5.980-00 "Hygienic requirements for surface water protection". (in Russian).
12. SanPiN 2.1.4.1074-01. Drinking water. Hygienic requirements to water quality of centralized drinking water supply systems. Quality control. (in Russian).
13. SanPiN 2.1.4.1175-02 Hygienic requirements for water quality of non-centralized water supply. Sanitary protection of sources. (in Russian).
14. Smith A. F. M. and Spiegelhalter D. J. Bays factors and choice criteria for linear models, *Journal of the Royal Statistical Society*. 1980; 42(2): 213-220.
15. Pearl, J. Graphs, causality, and structural equation models. *Sociological Methods and Research*. 1998; 27: 226-284.
16. ISO 5725-1: 1994, Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions.
17. ISO 5725-2: 1994, Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.
18. ISO 5725-3: 1994, Accuracy (trueness and precision) of measurement methods and results - Part 3: Intermediate measures of the precision of a standard measurement method.
19. ISO 5725-4: 1994, Accuracy (trueness and precision) of measurement methods and results - Part 4: Basic methods for the determination of the trueness of a standard measurement method.
20. ISO 5725-5: 1998, Accuracy (trueness and precision) of measurement methods and results - Part 5: Alternative methods for the determination of the precision of a standard measurement method.
21. ISO 5725-6: 1994, Accuracy (trueness and precision) of measurement methods and results - Part 6: Use in practice of accuracy values.
22. Bikbulatov E. S., Stepanova I. E. Harrington's desirability function for natural water quality assessment. *Russian Journal of General Chemistry*. 2011; 81(13): 2694-2704.
23. Korolyuk V.S., Portenko N.I., Skorokhod A.V., Turbin A.F. *Handbook of probability theory and mathematical statistics*. Moscow: Science, 1985. 640 p. (in Russian).
24. Rife G., Schleifer R. *Applied statistical decision theory*. M.: Statistics, 1977. 360 p. (in Russian).
25. GOST R 54500.3-2011/ISO/IEC Guide 98-3:2008 Uncertainty of measurement. Part 3. The guide to the expression of uncertainty in measurement/ (in Russian).
26. GOST R ISO 21748-2012. Statistical technology. Guidance on the use of estimates of repeatability, repeatability, and accuracy in assessing measurement uncertainty. (in Russian).
27. Shweta Tyagi, Bhavtosh Sharma, Prashant Singh, Rajendra Dobhal. Water Quality Assessment in Terms of Water Quality Index. *American Journal of Water Resources*. 2013; 1(3): 34-38.
28. Harashit Kumar Mandal. Influence of Wastewater PH on Turbidity. *International Journal of Environmental Research and Development*. 2014; 4(2):105-114.
29. Taguchi Genichi, Chowdhury Subir. *Taguchi's Quality Engineering Handbook*. 1696 p. 2004.
30. Kryukov S.P., Bodrunov S.D., Alexander L.N. and other. *Methods of analysis and assessment of risks in the tasks of safety management of complex technical systems*. SPb.: Aerospace equipment Corporation, 2007. 460 p. (in Russian).